Accepted Manuscript

Title: Adult fly age estimations using cuticular hydrocarbons and Artificial Neural Networks in forensically important Calliphoridae species

Authors: Hannah E. Moore, John B. Butcher, Charles R. Day, Falko P. Drijfhout

PII: S0379-0738(17)30402-4
DOI: https://doi.org/10.1016/j.forsciint.2017.10.001
Reference: FSI 8999

To appear in: *FSI*

Received date: 24-4-2017
Revised date: 28-9-2017
Accepted date: 2-10-2017

Please cite this article as: Hannah E.Moore, John B.Butcher, Charles R.Day, Falko P.Drijfhout, Adult fly age estimations using cuticular hydrocarbons and Artificial Neural Networks in forensically important Calliphoridae species, Forensic Science International https://doi.org/10.1016/j.forsciint.2017.10.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Forensic Science International, Volume 280, November 2017, pp. 233-244
DOI:10.1016/j.forsciint.2017.10.001

Adult fly age estimations using cuticular hydrocarbons and Artificial Neural Networks in forensically important Calliphoridae species

Hannah E. Moore[a*], John B Butcher[b], Charles R. Day[b], Falko P. Drijfhout[c]

[a*] *Corresponding author: Cranfield Forensic Institute, Cranfield University, Shrivenham, Swindon, SN6 8LA, h.e.moore@cranfield.ac.uk, 00441793 785257*

[b]*School of Computing and Mathematics, Keele University, Staffordshire, ST5 5BG*

Chemical Ecology Group, School of Chemical and Physical Sciences, Keele University, Staffordshire, ST5 5BG, UK

Highlights

- Promising technique using cuitcular hydrocarbons to age forensically important adult blowflies
- This technique provides accurate ageing to five time frames between 1 and 30 days old.
- Utilises PCA and ANN to visualise and separate out the ages of the blowflies
- Potential indoor crime scenes where adult flies are unable to escape.

Abstract

Blowflies (Diptera: Calliphoridae) are forensically important as they are known to be one of the first to colonise human remains. The larval stage is typically used to assist a forensic entomologists with adult flies rarely used as they are difficult to age because they remain morphologically similar once they have gone through the initial transformation upon hatching. However, being able to age them is of interest and importance within the field. This study examined the cuticular hydrocarbons (CHC) of Diptera: Calliphoridae species *Lucilia sericata*, *Calliphora vicina* and *Calliphora vomitoria*. The CHSs

Published by Elsevier. This is the Author Accepted Manuscript issued with: Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 4.0). The final published version (version of record) is available online at DOI:10.1016/j.forsciint.2017.10.001. Please refer to any applicable publisher terms of use.

were extracted from the cuticles of adult flies and analysed using Gas Chromatography-Mass Spectrometry (GC-MS). The chemical profiles were examined for the two Calliphora species at intervals of day 1, 5, 10, 20 and 30 and up to day 10 for *L. sericata*. The results show significant chemical changes occurring between the immature and mature adult flies over the extraction period examined in this study. With the aid of a Principal Component Analysis (PCA) and Artificial Neural Networks (ANN), samples were seen to cluster, allowing for the age to be established within the aforementioned time frames. The use of ANNs allowed for the automatic classification of novel samples with very good performance. This was a proof of concept study, which developed a method allowing to age post-emergence adults by using their chemical profiles.

Keywords: adult blowflies, artificial neural networks, cuticular hydrocarbons, post mortem interval, principal component analysis.

Introduction

Blowflies are the first wave of insects to arrive at a corpse and therefore are of significant forensic importance. They provide a substantial amount of evidence regarding the time of death, location of death and toxicology information [1,2]. When a forensic entomologist is presented with an indoor crime scene where adult flies have accumulated by, for example a window, it would be highly beneficial if the age of the flies could be established. [3].

Recent studies have examined the use of new developments in techniques such as DNA [4-6] and Cuticular Hydrocarbons (CHC) [3][7-13] to identify and age various life stages of forensically important blowflies and are being incorporated into the field of forensic entomology. These techniques are being applied with the overall aim of strengthening identification and ageing difficulties encountered by forensic entomologists and taxonomists. Studies have been published using hydrocarbons,

ACCEPTED MANUSCRIPT

found on the cuticles of all insects to successfully identify and age the larval life stages [8,9].

The cuticle of insects is covered by a layer of epicuticular waxes, consisting of hydrocarbons, fatty acids, waxes and alcohols [14]. The main purpose of this cuticular lipid layer is to prevent desiccation as well as penetration of micro-organisms [15]. The hydrocarbons are known to be stable [11][16,17] and are found to be a very good means of identifying insects as the hydrocarbon profiles are known to be species specific [14,15][18-21]. This study applied the same techniques as presented by Pechal and co-workers [3] to examine the chemical profiles of adult flies from three forensically important species commonly found in the UK, *Lucilia sericata*, *Calliphora vicina* and *Calliphora vomitoria*. Once a fly has emerged from the puparia and taken on the appearance of a mature adult fly, there are no significant morphological changes that occur in order for their age to be established, making accurate age estimation difficult. Cuticular hydrocarbons have the potential to be a reliable age indicator as they are known to be stable throughout the development of the blowflies life cycle [3][8,9].

Principal Component Analysis (PCA) is often used on complex datasets [18] to interpret the chromatograms in order to ease visualization of any trends that maybe present. For example, PCA has already been used by the authors to age the life stages of *Lucilia sericata*, *Calliphora vicina* and *Calliphora vomitoria* [3,23]. The results showed that in most cases, larvae can be aged to the day, as appose to the instar.

Artificial Neural Networks (ANNs) are a well-studied and utilized machine learning approach [24]. Taking inspiration from the learning mechanisms observed in the biological brain, their appeal stems from their ability to learn characteristics of a dataset and classify new instances that were previously unseen through the

ACCEPTED MANUSCRIPT

modification of their weighted connections. This data can be from real-world domains, therefore noisy and in some cases non-linear, two data characteristics that ANNs are well-suited to analyse. These strengths make them appealing for use in domains where new data occurs often and requires fast and accurate analysis, such as PMI estimation, but also many other domains [13][25-27].

Previous work showed [13][28] a particular type of ANN, known as the Self-Organising Map (SOM) [29], to offer good generalization capabilities when analyzing hydrocarbons collected from larvae of the three same species analysed in this current work. Using ANNs has the added advantage of automated classification once training is complete, a characteristic that is not offered by PCA or some other analysis approaches.

A SOM is perhaps the most widely studied unsupervised ANN which consists of an input and output layer containing artificial neurons, each of which is connected to every other neuron in its neighboring layer (e.g. an input neuron is connected to every neuron in the output layer). The output layer reveals the clustering of the input data based on its underlying characteristics which occurs during training. Once training is complete input patterns that are similar in their characteristics should cluster in similar regions of the output layer, while input patterns that are different do not. By projecting the input data onto a high dimensional state space, good classification can be achieved using ANNs that may be not possible or visibly obvious using other approaches such as PCA, where only two or three dimensions are visualized.

Unsupervised training does not require class labels (e.g. one-day old adult fly) to be assigned to input patterns, thereby removing this potentially laborious task (especially in the case of large datasets). The clustering of input patterns can also lead to insights into the data that was not apparent prior to training through the SOM's topological

ACCEPTED MANUSCRIPT

ordering capabilities [29]. For example, an input pattern that is projected between two clusters on the output layer may be a mix of these two classes. The added insights into datasets provided by the SOM has been found by others when analyzing GC-MS data [30] and could be useful in many more domains within analytical chemistry where other approaches are not suitable [31].

The hydrocarbons were analysed by Gas Chromatography – Mass Spectrometry (GC-MS) and to aid visualisation of the results, Principal Component Analysis (PCA) was used to show any datapoints that may cluster based on similar characteristics, while ANNs were applied to the datasets with the aim of providing an automated classifier that can age novel adult flies based on previous seen examples.

The overall aim of this preliminary study was to determine whether time dependent chemical changes occur within the adult profiles of *L. sericata*, *C. vicina* and *C. vomitoria* over a period of 30 days, allowing for the age of the adult flies to be established.

Materials and Methods

Insect materials

A colony of *Calliphora vicina* and *Calliphora vomitoria*, (geographical origin, University of Birmingham campus, UK), kindly supplied by the Scott Hayward's research group at the University of Birmingham and *Lucilia sericata* (geographical origin, Haywards Heath, West Sussex, UK), kindly supplied by the Natural History Museum in London, were reared in the laboratory and maintained in separate rearing cages under standard environmental conditions (22±1 ^{o}C). They were fed with sugar, water and milk powder with pig's liver as an egg laying medium [33]. After one hour of placing the meat into the rearing cage, the meat with the laid eggs were removed

ACCEPTED MANUSCRIPT

and placed into a box for rearing, with the larvae being fed regularly until post-feeding. Once pupation started, all pupae that had pupated within a 2 hour time window (approximately 200) were removed and placed into a separate rearing cage. When the adult flies emerged (approximately 85%) any unhatched pupae were removed to ensure all adult flies emerged on the same day, within two hours of each other. This was then the sample pool used to extract on days 1, 5, 10, 20 and 30 for *C. vicina* and *C. vomitoria* and days 1, 5 and 10 for *L. sericata*.

Sample Preparation

Liquid extraction with hexane was used to extract the hydrocarbons. For each sample (n=10) a single adult fly was used which provided a sufficient concentration for the GC-MS to detect the hydrocarbons. The flies were submerged in hexane and. left for 10 to 15 minutes. The hexane was removed and transferred to a clean GC vial and left until completely dry. The extracts were re-dissolved in 150 µl of hexane and a 2 µl aliquot was injected into the GC-MS via the autosampler [32].

Chemical Analysis: Gas Chromatography – Mass Spectrometry

Chemical analysis of all extracts was carried out on an Agilent Technologies 6890N Network GC with a split/splitless injector at 250 ^{o}C, a Restek Rxi-1MS capillary column (30m x 0.25 mm ID, 0.25µm film thickness) coupled to an Agilent 5973 Network Mass Selective Detector. The GC was coupled to a computer and data processed with Agilent Chemstation software. Elution was carried out with helium at 1mL/min. The oven temperature was programmed to be held at 50 ^{o}C for 2 minutes then ramped to 200 ^{o}C at 25 ^{o}C/min, then from 200 ^{o}C to 260 ^{o}C at 3^{o}C/min and finally from 260 ^{o}C to 320 ^{o}C at 20 ^{o}C/min where it was held for 2 minutes. The mass

ACCEPTED MANUSCRIPT

spectrometer was operated in Electron Ionisation mode at 70 eV, scanning from 40 – 500 amu at 1.5 scans s^{-1}. Hydrocarbons were identified using a library search (NIST08), the diagnostic fragmented ions and the Kovats indices [1].

Statistical analysis: Principal Component Analysis

The calculations were carried out using a multivariate add-in to Microsoft Excel, which was written by Tom Thurston using an original development by Les Erskine [33]. As standard practice, ten replicates at each extraction age were taken on all occasions (*n=10*). Six principal components were used as they described 100% of the variation within the dataset. Our methods of Principal Component Analysis (PCA) have been reported previously [23,34].

Artificial Neural Networks

The same training and testing approaches detailed in Butcher et al. [13] and Moore et al. [28] were used to train a SOM to classify previously unseen hydrocarbon profiles of adult *L. sericata*, *C. vicina* and *C. vomitoria* flies. Very briefly, training of a SOM involves presenting every input pattern to the input layer where the weighted connection between the input neuron and the output neuron whose activation closely matches the input pattern (the winning neuron) are modified by a standard Hebbian learning rule [29]. This process is repeated until well-formed clusters appear in the output layer where similar patterns cluster in close proximity to each other in the output layer. Readers are referred to Kohonen [29] and Butcher et al. [13] for more details on SOM training.

The data processing used in this study is the same as described previously (Butcher et al. [13] and Moore et al. [28]). First, the data was preprocessed using PCA to reduce

$$v_n = \frac{v - \min(v)}{\max(v) - \min(v)} \times (U_{lim} - L_{lim}) - L_{lim}$$

ACCEPTED MANUSCRIPT

the dimensionality of the dataset, where the first 6 principal components were used. The data was then normalized between the range +1 and -1 by:

where v_n is the normalised data, v is the original data, $max(v)$ and $min(v)$ are the maximum and minimum data values respectively and $_{Ulim}$ and $_{Llim}$ are the desired upper and lower limits of the normalised data which in this study are set to +1 and -1 respectively. Training data input to the SOM contained the average of five hydrocarbon profiles for each species' adult flies. The generalization capabilities of the SOM were then assessed using two approaches:

1) Using an average of five hydrocarbon samples as the unseen test dataset.
2) Using the remaining unseen individual hydrocarbon samples as the unseen test dataset.

Training and testing of each SOM was conducted using 10-fold cross-validation where the average test performance and standard deviation was calculated. During each fold a unique random subset of each day's hydrocarbon profiles was chosen for the training and testing datasets (note: the same subsets were used for both testing approaches outlined above). The size of the output layer was systematically evaluated to determine the size that provided the best performance as determined by the clusters formed upon completion of training. During training, two important SOM training parameters, the learning rate and neighborhood size were updated following the same approach outlined in Day et al. [35].

Results

GC-MS analysis

ACCEPTED MANUSCRIPT

A total of 130 individuals were analysed; 50 *C. vicina*, 50 *C. vomitoria* and 30 *L. sericata*. GC-MS successfully separated and identified 48 compounds for *L. sericata*, 60 compounds for *C. vicina* and 55 compounds for *C. vomitoria*. The profiles contained mainly saturated hydrocarbons, which were observed in all three species with chain lengths ranging from C21 to C31 carbons. Alkenes and mono-, di-, tri- and tetra-methyl branched hydrocarbons were present in the profiles of the blowflies, with varying chain lengths and compositions, allowing for distinguishes to be made between the three species.

Figures 1-3 show the GC chromatograms of a single adult fly sample from day 1, 5 and 10 for *L. sericata* (Figure 1) and days 1, 5, 10, 20 and 30 for *C. vicina* (Figure 2) and *C. vomitoria* (Figure 3). Chemical distinctions can be made between the different fly ages, for example peak 3 (tricosane) gives a clear distinction between younger and older adult flies (days 1 and 5).

There are a number of characteristics that could be utilised as potential age indicators, for example, the number of *n*-alkanes increases for *L. sericata* and the number of methyl branched compounds decreases with age for *L. sericata* and *C. vicina*.

Table 1 lists the compounds extracted from *L.sericata* over the 10 days.

When looking at the proportions of the hydrocarbons present in the profile of *L. sericata*, more characteristic trends are observed. In general, the relative percentage of the lower molecular weight compounds (C21 to C27) increase with age, whilst the higher molecular weight compounds generally decrease with age (with the exception of a few). Pentacosane (compound 8) and heptacosane (compound 18) are seen to significantly increase with time. The alkenes also make good age indicators as four of

ACCEPTED MANUSCRIPT

them increase significantly with age (tricosene, heptacosene, nonacosene and tritriacontene).

Tables 2 and 3 list the compounds extracted from *C. vicina* and *C. vomitoria* over a 30 day period.

The concentrations of some of the compounds in the profile of *C. vicina* change systematically with time, which can also be a good indicator of age. The concentration of tricosane increases significantly between days 1 and 5, then further decreases with time and is not observed in a concentration above 0.5% in the final day of extractions (day 30). Pentacosane is present for all 5 extraction days and increases with time. Octacosane is only present in an adequate proportion in days 1, 5 and 10 but can be seen to decrease with time over the three ages. Nonacosane is at its most abundant in day 1, before remaining relatively stable across the other 4 extraction days.

In the chemical profiles of *C. vomitoria*, day 1 has a substantially large number of compounds specific to that day (31 in total – compounds 19-25, 28, 30-39, 41-42, 44-45 and 47-55), which mainly consist of long chain length methyl branched compounds (C27 up to C33). Day 5 has two compounds specific to that day, 2-Methyloctacosane (compound 26) and hentriacontane (compound 43). The profiles of days 10 and 20 have no compounds specific to those individual days. Day 30 has three compounds that are not observed in any of the other extraction days. These compounds are a tricosene isomer (compound 4), tetracosane (compound 9) and a heptacosene isomer (compound 17).

ACCEPTED MANUSCRIPT

Principal Component Analysis

L. sericata: All 48 of the compounds extracted from the adult flies profile were used for PCA analysis, of which 63% were methyl branched, 23% were alkenes and 15% of the hydrocarbons were *n*-alkanes. Principal component analysis was carried out using 6 principal components (same applies for the *Calliphora* species). These six components described 99.0% of the variation within the data set with the first two (PC1 and PC2) comprising 65.5% and 24.3% respectively.

Figure 4 shows the PCA plot of PC3 vs PC2 for data gathered from day 1, 5, 10 of adult fly extractions for *L. sericata*.

C. vicina: All 60 compounds extracted from the adult flies profile were used for PCA analysis, of which 70% were methyl branched, 18% were alkenes and 12% of the hydrocarbons were *n*-alkanes. The six components described 97.7% of the variation within the data set with the first three (PC1, PC2 and PC3) comprising 71.0%, 15.3% and 6.2%. Figure 5 shows the PCA plot of PC3 against PC2 for data gathered from day 1, 5, 10, 20 and 30 of adult fly extractions for *C. vicina*.

C. vomitoria: 55 compounds extracted from the adult flies profile were used for PCA analysis, of which 70% were methyl branched, 15% were alkenes and 15% of the hydrocarbons were *n*-alkanes. The six components described 99.2% of the variation within the data set with the first two (PC1 and PC2) comprising 64.1% and 19.9%. Figure 6 shows the PCA plot of PC3 against PC2 for data gathered from day 1, 5, 10, 20 and 30 of adult fly extractions for *C. vomitoria*.

From the PCA plots, it is evident to see clustering groups which represent the different ages of the flies. For *L. sericata* the three extraction days are all clustered in individual groups within the plot. For the plot of *C. vicina,* the flies can be aged to the

ACCEPTED MANUSCRIPT

time frames examined with the exception of the mature ages of day 20 and 30 which group together. *C. vomitoria* flies can be aged up to 10 days old but after this age they cannot be accurately separated using PCA for visualization.

Neural Network Analysis

Table 4 shows the generalisation performance of the SOM across the three datasets when presented with the average of 5 samples as the test data (top row) or individual samples (bottom row) after training using the average of 5 samples was complete. These results show that the SOM offers very good classification performance when using the average of 5 samples compared to using individual samples for testing. This to be expected due to the reduction in variability during training, which was also the case when analyzing larvae of the same species [13][28].

Perfect classification is achieved for *L. sericata* and almost near perfect classification for *C. vomitoria* is achieved. The PC plots in Figure 4 show that *L. sericata* has distinct PC clusters, while the clusters for *C. vomitoria* (Figure 6) show good separation with the exception of the older flies. The clustering of *C. vicina* (Figure 5) is less well defined as older flies group together. These less well-defined PC clusters could explain the reduced performance of the SOM for these two datasets where samples of different ages may share similar characteristics, making accurate classification more difficult. This is shown in the bottom two confusion matrices of each SOM for *C. vicina* and *C. vomitoria* in Table 5, where for *C. vicina* 50% of 20 day old adults are incorrectly classified as 30 days old and for *C. vomitoria* one 10 day old adult is classified as a 20 day old adult. Despite these errors, the SOM is able to offer good classification performance, particularly in the case of classifying *C.*

ACCEPTED MANUSCRIPT

vomitoria adults that are older than 10 days old with only one misclassification. Upon visual inspection of the PC plots for *C. vomitoria* in Figure 6 this good performance may be surprising as plotting PC2 and PC3 reveals very tight clustering for 10-30 day old adults that look inseparable. However, the near perfect test performance of the SOM reveals the power of ANNs, whereby using all 6 PCs and projecting the input data onto a higher dimensional state space, better clustering and classification can be achieved when compared to plotting two PCs for data visualization.

Discussion

The only morphological changes that occur within adult flies happen with in the first few hours after a fly has emerged. When an adult emerges from the puparial cases, it has an unusual appearance. There is a protruding region of the flies head, called the ptilinum, which becomes inflated with hemolymph and enables them to push their way out of the cases [36][1]. This retracts after a few hours, forming the ptilinal suture of the head. The wings of the fly are also crumpled and the bodies are brown/grey in colour. However, once the ptilinum sinks back into the facial structure, the flies wings are fully formed and it gains its colour, there are no other morphological characteristics to determine the age [33]. Previous studies have investigated other means of ageing the mature adult fly by examining the pteridine levels [37, 38], cuticular band counting [39] and rates of ovarian development [3]. Hydrocarbon analysis of insects such as grasshoppers [40] and cockroaches [41] have been carried out to establish the age and sex. Hydrocarbon studies have also been carried out on Diptera [21][42-44] and more specifically Calliphoridae [7], but few have been in relation to its importance in forensic entomology [8]. Trabalon and co-workers [7] examined the CHCs of the adult fly species *C. vomitoria* in relation to age

ACCEPTED MANUSCRIPT

and sex. The results they presented were very promising and they observed changes in the epicuticular hydrocarbon content and composition of both sexes of the species. They were able to observe differences within the CHC profiles allowing for a distinction between young adults (3-6 hr) to older ones (24-120 hr). They also noted a significant decrease in the long chain methyl branched compounds as the fly aged which was true of the same species (*C. vomitoria* in this species, which can be seen from day 5 onwards (Table 3). However, they only examined flies up to 8 days old. In indoor scenes where the flies cannot escape, being able to age older flies would be advantageous. Roux and colleagues [8] also carried out a complete ontogenetic study of *C. vomitoria*, *C. vicina* and *P. terraenove* ranging from eggs to adult flies. They also observed distinguishable profiles from young (aged < 24 h) and old (aged >24 h) adult flies. However, like Trabalon [7], they only examined flies up to 8 days old. Results presented in this study are in agreement with Trabalon and Roux in that there is a significant difference between young (day 1) adult flies and older adults (> 5 days) but the results in this paper go further by accurately aging adult flies up to the 30 day time frame.

The chemical profile of the day 1 adult flies for all three species are very distinct as they are dominated by mono-, di- and tri-methyl hydrocarbons. As the adult fly ages, the shorter chain length methyl branched compounds start to decrease in their relative proportions. The more complex hydrocarbons observed in the young adult flies could originate from the larvae/pupae stage, and as the fly ages become less dominant. This could be linked with the fact that the mature fly requires less flexibility in its cuticle[14][45].

The data presented in this study was a continuation of that carried out by Pechal et al (2014) [3] where cuticular hydrocarbons were used to age *Chrysomya rufifacies*

ACCEPTED MANUSCRIPT

and *Cochliomyia macellaria*. As with the results presented by Pechal et al. [3], the results within this study also show distinguishable differences observed within the chemical profiles. For all three species, ageing can be established to the day between days 1 and 5, and 10 for *C. vicina* and *L. sericata*. The mature ages of days 20 and 30 then group with day 10 for the profiles of *C. vomitoria*, whereas *C. vicina* has a 4th cluster group containing days 20 and 30 when visualizing the data in PC space using PCA.

Despite the relatively small dataset, a self-organising map (SOM) was able to classify each of the adult fly ages with very good performance, achieving 100% classification accuracy in the case of the *L. sericata*, and at least 70% classification accuracy for *C. vicina* and *C. vomitoria*. The SOM was also able to offer good classification for the older flies (10+ days old) across all species, something that was not possible to visualize when plotting two principal components: a further advantage of applying SOMs for the automated classification of hydrocarbon profiles for PMI_{min} estimation. A larger dataset using samples collected in the field would be interesting future work, where it is hypothesized that SOMs or similar ANN approaches would be well-suited. The use of a SOM to cluster the data and then automatically classify new instances of novel adult flies has great potential to provide forensic scientists with an automated tool for ageing post-emergence flies.

Further investigation into the cuticular hydrocarbons of adult flies with the aim to achieve smaller ageing time frames from 5 day intervals, as presented in this study, to just 1 day would be advantageous. The effect of temperature and humidity should also be examined to test the stability of the hydrocarbons and see if this has an effect on the chemical profiles or weather an ageing model can still be applied.

ACCEPTED MANUSCRIPT

The results presented in this preliminary study show great potential for ageing a life stage that is currently not utilized within the field of forensic entomology due to the fact that there is currently no other means to do so. Further work at different temperature and humidity levels is required to ensure the stability of the chemical profiles in different environments, which will require similar statistical and ANN analysis that has been shown in this present work.

ACCEPTED MANUSCRIPT

References

[1] Benecke, M. (2001) A brief history of forensic entomology. *Forensic Science International*, **120**, 2-14.

[2] Amendt, J., Krettek, R., and Zehner, R. (2004) Forensic entomology. *Naturwissenschaften*. **91**, 51-56.

[3] Pechal, J., Moore, H., Drijfhout, F.P., and Benbow, M.E. (2014) Hydrocarbon profiles throughout adult Calliphoridae aging: A promising tool for forensic entomology. *Forensic Science International*, 245, 65-71.

[4] Tarone, A., and Faran, D. (2011) Gene expression during blow fly development: Improving the precision of age estimates in forensic entomology. *Journal of Forensic Sciences*, **56**, S112-S122.

[5] Harvey, M., Dadour, I., and Guadieri, S. (2003) Mitochondrial DNA cytochrome oxidase I gene: potential for distinction between imature stages of some forensically important fly species (Diptera) in western Australia. *Forensic Science International*, **131**, 134-139.

[6] Boehme, P., Spahn, P., Amendt, J., and Zehner, R. (2013) Differential gene expression during metamorphosis: a promising approach for age estimation of forensically important *Calliphora vicina* pupae (Diptera: Calliphoridae). International Journal of Legal Medicine, 127, 243-249.

[7] Trabalon, M., Campan, M., Clement, J.L., Lange, C., and Miquel, M.T. (1992) Cuticular hydrocarbons of *Calliphora vomitoria* (Diptera): Relation to age and sex. *General and Comparative Endocrinology*, **85**, 208–216.

[8] Roux, O., Gers, C., and Legal, L. (2008) Ontogenetic study of three Calliphoridae of forensic importance through cuticular hydrocarbon analysis. *Medical and Veterinary Entomology*, **22**, 309–317.

[9] Zhu, G.H., Ye, G.Y., Hu, C., Xu, X.H., and Li, K. (2006) Development changes of cuticular hydrocarbons in *Chrysomya rufifacies* larvae: potential for determining larval age. *Medical and Veterinary Entomology*, **20**, 438–444.

ACCEPTED MANUSCRIPT

[10] Zhu, G., Xu, X., Yu, X., Zhang, Y., and Ye, J. (2007) Puparial case hydrocarbons of *Chrysomya megacephala* as an indicator of the postmortem interval. *Forensic Science International*, **169**, 1-5.

[11] Ye, G., Li, K., Zhu, J., Zhu, G., and Hu, C. (2007) Cuticular Hydrocarbon Composition in Pupal Exuviae for Taxonomic Differentiation of Six Necrophagous Flies. Neurobiology, *Journal of Medical Entomology*, **44**, 450-456

[12] Bernhard, V., Pogoda, W., Verhoff, M., Toeness, S., and Amendt, J. (2017) Estimating the age of the adult stages of the blow flies *Lucilia sericata* and *Calliphora vicina* (Diptera: Calliphoridae) by means of the cuticular hydrocarbon *n*-pentacosane. *Science and Justice*, doi.org/10.1016/j.scijus.2017.04.007

[13] Butcher, J.B., Moore, H.E., Day, C.D., Adam, C.D., and Drijfhout, F.P. (2013) Artificial neural network analysis of hydrocarbon profiles for the ageing of *Lucilia sericata* for post mortem interval estimation. *Forensic Science International*, **232**, 25-31.

[14] Gibbs, A.G. (1998) Water-Proofing Properties of Cuticular Lipids. *American Zoologist*, **38**, 471–482.

[15] Everaerts, C., Farine, J.P., and Brossut, R. (1997) Changes of species specific cuticular hydrocarbon profiles in the cockroaches *Nauphoeta cinerea* and *Leucophaea maderae* reared in heterospecific groups. *Behavioral Ecology and Sociobiology*, **85**, 145–150.

[16] Braga, M., Pinto, Z., Queiroza, M.C., Matsumotod, N., Blomquist, G. (2013) Cuticular hydrocarbons as a tool for the identification of insect species: Puparial cases from Sarcophagidae. *Acta Tropica*, **128**, 479-485.

[17] Steiner, S., Hermann, N., Ruther, J., (2006) Characterization of a Female-Produced Courtship Pheromone in the Parasitoid *Nasonia vitripennis*. *Journal of chemical ecology*, **32**, 1687-1702.

[18] Martin, S., and Drijfhout, F. (2009) How reliable is the analysis of complex cuticular hydrocarbon profiles by multivariate statistical methods? *Journal of Chemical Ecology*, **35**, 375–382.

ACCEPTED MANUSCRIPT

[19] Lavine, B.K., and Vora, M.N. (2005) Identification of Africanized honeybees, *Journal of Chromatography. A*, **1096**, 69–75.

[20] Lockey, K. (1992) Insect hydrocarbon chemotaxonomy: Cuticular hydrocarbons of adult and larval *epiphysa* species blanchard and adult *Onymacris unguicularis* (HAAG) (tenebrionidae: Coleoptera). *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, **102**, 451–470.

[21] Urech, R., Brown, G.W., Moore, C.J., and Green, P.E. (2005) Cuticular hydrocarbons of buffalo fly, *Haematobia exigua*, and chemotaxonomic differentiation from horn fly, *H. irritans*. *Journal of Chemical Ecology*, **31**, 2451–61.

[22] Guillem, R.M., Drijfhout, F.P., and Martin, S.J. (2012) Using chemo-taxonomy of host ants to help conserve the large blue butterfly. *Biological Conservation*, **148**, 39–43.

[23] Moore, H.E., Adam, C.D., and Drijfhout, F.P. (2012) Potential Use of Hydrocarbons for Aging *Lucilia sericata* Blowfly Larvae to Establish the Postmortem Interval. *Journal of forensic sciences*, **58**, 404–412.

[24] Haykin, S. (2009) *Neural Networks and Learning Machines*, Pearson Education, New Jersey, USA.

[25] Blanconi, A., Von Zuben, C., Serapião, A., and Govone, J. (2010) Artificial neural networks: a novel approach to analysing the nutritional ecology of a blowfly species, Chrysomya megacephala. *Journal of Insect Science*, **10**, 1-18.

[26] Abdullahi, S., Mathias S., and Pretzsch. H. (2017) An unsupervised two-stage clustering approach for forest structure classification based on X-band InSAR data—A case study in complex temperate forest stands. *International Journal of Applied Earth Observation and Geoinformation*, **57**, 36-48.

[27] Fei, B.K., Eloff, J.L., Olivier, M.S., and Venter, H.S. (2006) The use of self-organising maps for anomalous behaviour detection in a digital investigation. *Forensic Science International*, **162**, 33-37.

ACCEPTED MANUSCRIPT

[28] Moore, H.E., Butcher, J.B., Adam, C.D., Day, C.D., and Drijfhout, F.P. (2016) Age estimation of Calliphora (Diptera: Calliphoridae) larvae using cuticular hydrocarbon analysis and Artificial Neural Networks. *Forensic Science International*, **268**, 81-91.

[29] Kohonen, T. (1990) The self-organising map. *Proceedings of the Institute of Electrical and Electronics Engineers*, **78**, 1464–1480.

[30] Steiner, F., Schlick-Steiner, B., Nikiforov, A., Kalb, R., Mistrik, R. (2002) Cuticular hydrocarbons of Tetramorium ants from central Europe: analysis of GC–MS data with self-organizing maps (SOM) and implications for systematics. *Journal of Chemical Ecology*, **28**, 2569–2584.

[31] Martin, S., Drijfhout, F. (2009) How Reliable is the Analysis of Complex Cuticular Hydrocarbon Profiles by Multivariate Statistical Methods? *Journal of Chemical Ecology*, **35**, 375–382.

[32] Moore, H. (2012) Analysis of cuticular hydrocarbons in forensically important blowflies using mass spectrometry and its application in post mortem interval estimations. PhD Thesis, Keele University.

[33] Brereton, R.G. (2003) *Chemometrics: Data analysis for the laboratory and chemical plant*. Wiley, Chichester UK.

[34] Adam, C., Sherratt, S., and Zholobenko, V. (2008) Classification and individualisation of black ballpoint pen inks using principal component analysis of UV-vis absorption spectra. *Forensic Science International*, **174**, 16-25.

[35] Day, C., Austin, J., Butcher, J., and Haycock, P. (2009) Element-specific determination of X-ray transmission signatures using neural networks. *Nondestructive Testing and Evaluation International*, **42**, 446-451.

[36] Moore, H., and Drijfhout, F. (2015) *Forensic Entomology: International Dimensions and Frontiers*. CRC Press, USA. pp 373.

[37] Wall, R., Langley, P., Stevens, J., and Clarke, G. (1990) Age-determination in the old-world screw-worm fly *Chrysomya bezziana* by pteridine fluorescence. *Journal of Insect Physiology*, **36**, 213–218.

ACCEPTED MANUSCRIPT

[38] Thomas, D., and Chen, A. (1989) Age determination in the adult screwworm (Diptera: Calliphoridae) by pteridine levels. *Journal of Economic Entomology*, **82**, 1140-1144.

[39] Bala, M., and Sharma, A. (2016) Review of some recent techniques of age determination of blow flies having forensic implications, *Egyptian Journal of Forensic Sciences*, **6**, 203–208.

[40] Tregenza, T., Buckley, S., Pritchard, V., and Butlin, R. (2000) Inter- and intra-population effects of sex and age on epicuticular composition of meadow grasshopper, *Chorthippus parallelus*. *Journal of Chemical Ecology*, **26**, 257–278.

[41] Brown, W.V., Rose, H.A., Lacey, M.J., and Wright, K. (2000) The cuticular hydrocarbons of the giant soil-burrowing cockroach *Macropanesthia rhinoceros* saussure (Blattodea: Blaberidae: Geoscapheinae): analysis with respect to age, sex and location. *Comparative Biochemistry and Physiology. Part B, Biochemistry and molecular biology*, 127, 261–77.

[42] Brown, W.V. Morton, R., and Spradbery, J.P. (1992) Cuticular hydrocarbons of the old world screw-worm fly, *Chrysomya bezziana* Villeneuve (Diptera: Calliphoridae). Chemical characterization and quantification by age and sex. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, **101**, 665–671.

[43] Mpuru, S., Blomquist, G.J., Schal, C., Roux, M., Kuenzli, M., Dusticier, G., Clément, J.L., and Bagnères, A.G. (2001) Effect of age and sex on the production of internal and external hydrocarbons and pheromones in the housefly, *Musca domestica*. *Insect Biochemistry and Molecular Biology*, **31**, 139–55.

[44] Hugo, L.E., Kay, B.H., Eaglesham, G.K., Holling, N., and Ryan, P.A. (2006) Investigation of cuticular hydrocarbons for determining the age and survivorship of Australasian mosquitoes. *The American Journal of Tropical Medicine and Hygiene*, **74**, 462–74.

[45] Morgan ED (2004) Biosynthesis in insects. The Royal Society of Chemistry Cambridge, UK

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

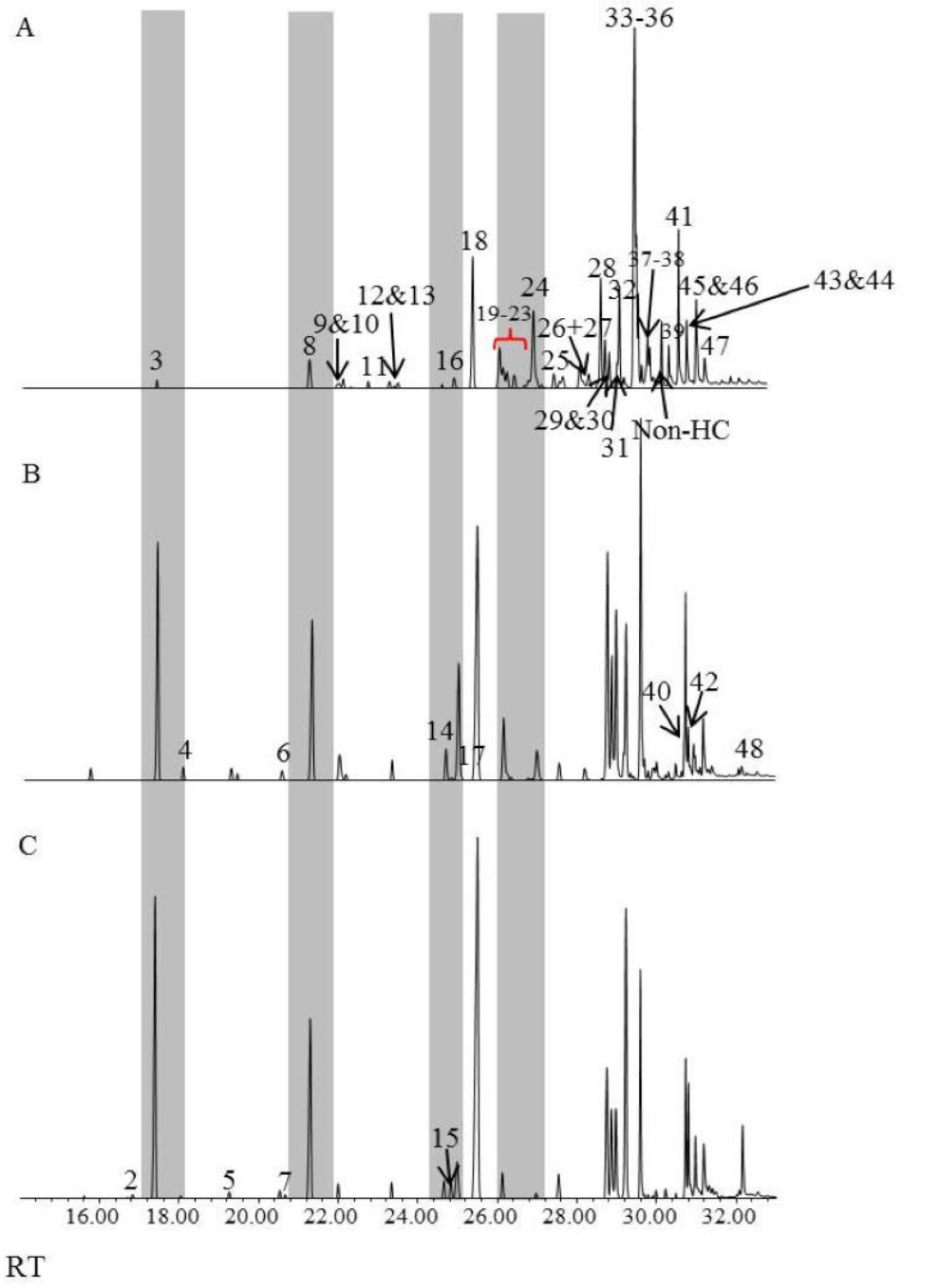

Figure 1: GC chromatograms of *L. sericata* adult fly at three different ages, A: Day 1, B: Day 5 and C: Day 10.

ACCEPTED MANUSCRIPT

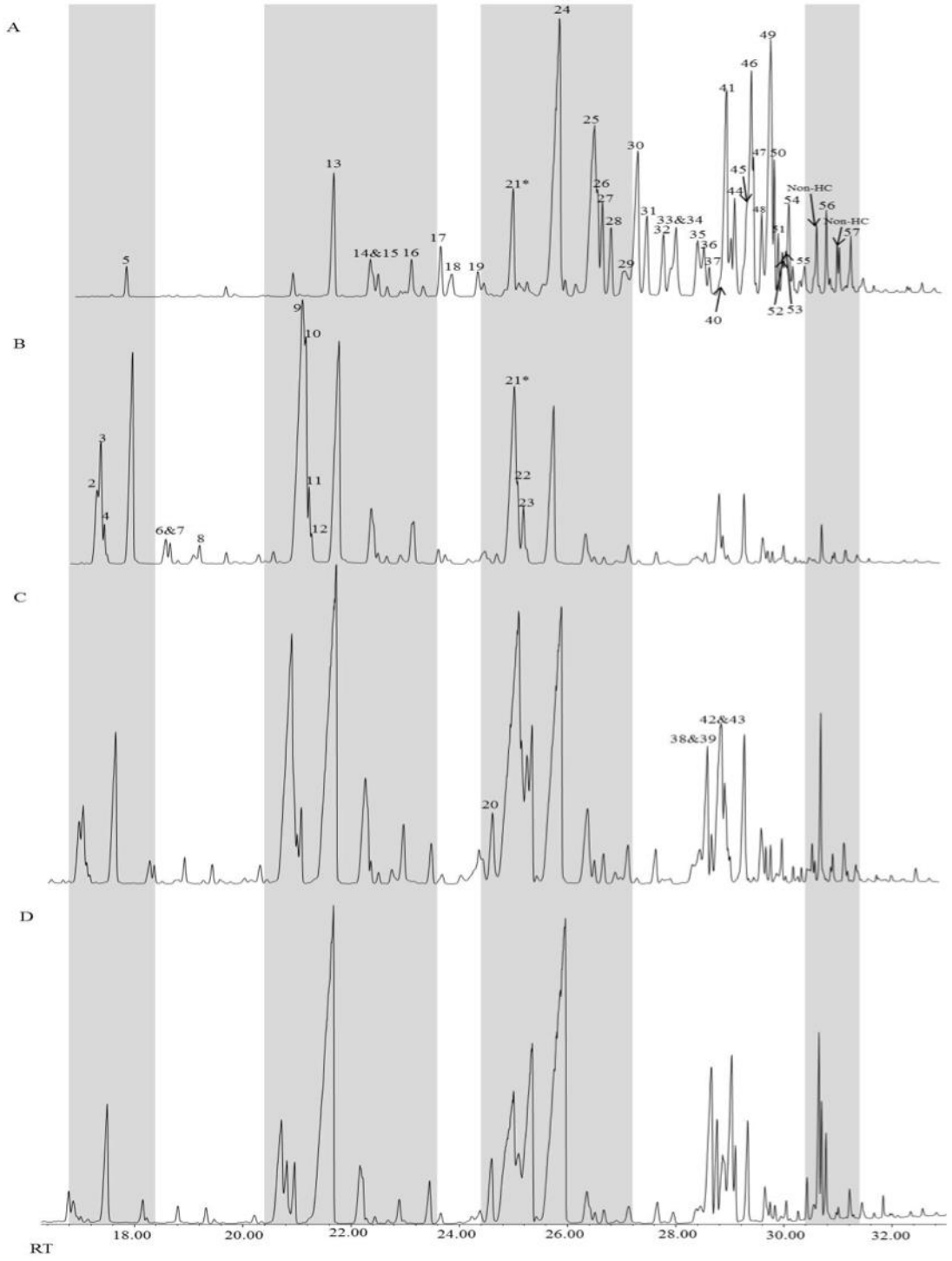

Figure 2: GC chromatograms of *C. vicina* adult flies at four different ages, A: Day 1, B: Day 5, C: Day 10 and D: Day 20. Shaded bars illustrate distinctive changes over time indicating specific areas of interest

ACCEPTED MANUSCRIPT

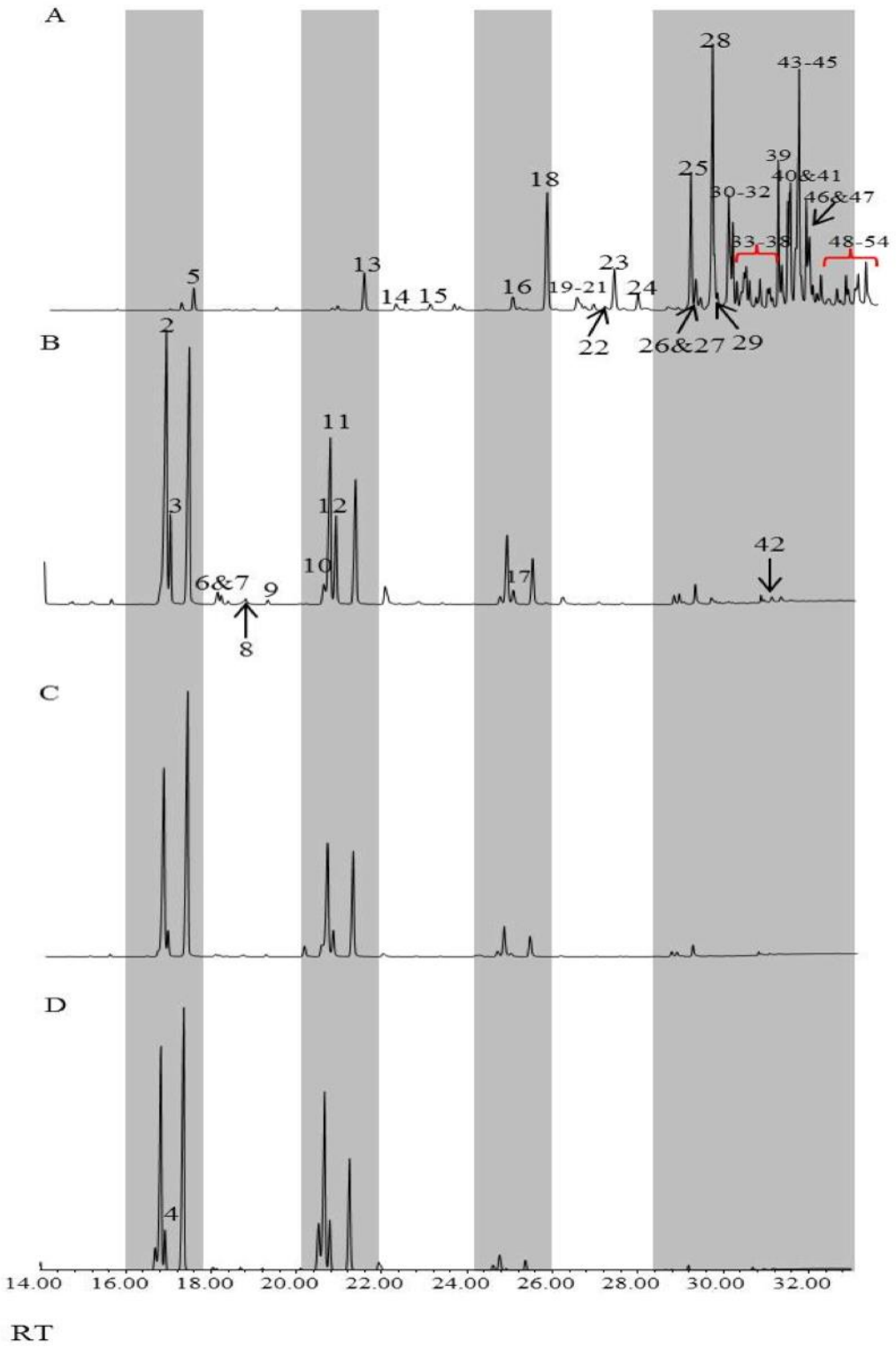

Figure 3: GC chromatograms of *C. vomitoria* adult flies at four different ages, A: Day 1, B: Day 5, C: Day 10 and D: Day 20. Shaded bars illustrate distinctive changes over time indicating specific areas of interest

ACCEPTED MANUSCRIPT

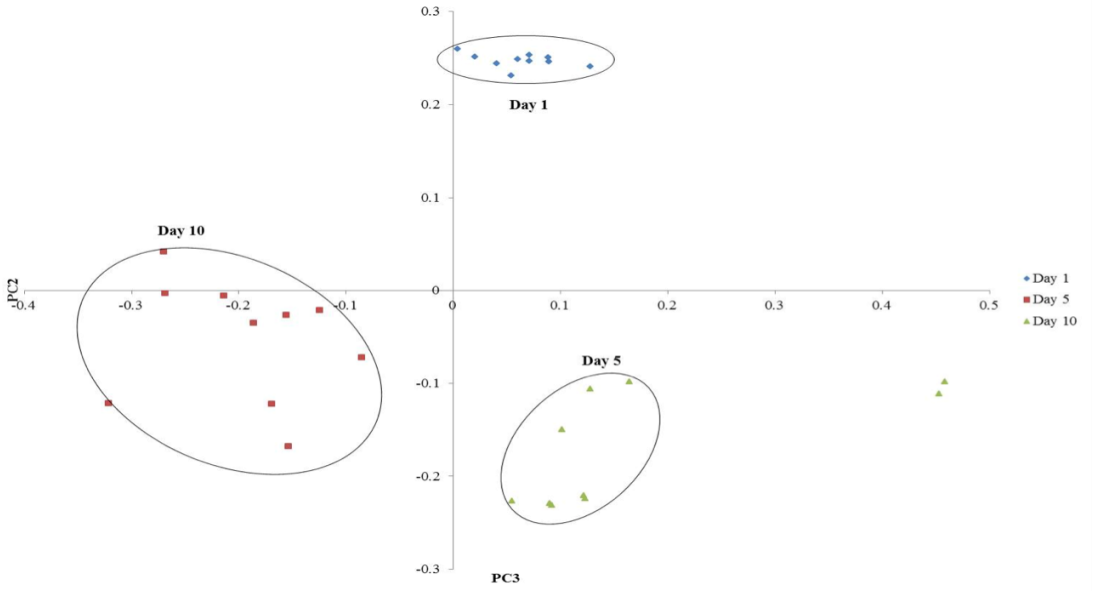

Figure 4: PCA plot showing PC3 against PC2 for *L. sericata* with clustering days circled

ACCEPTED MANUSCRIPT

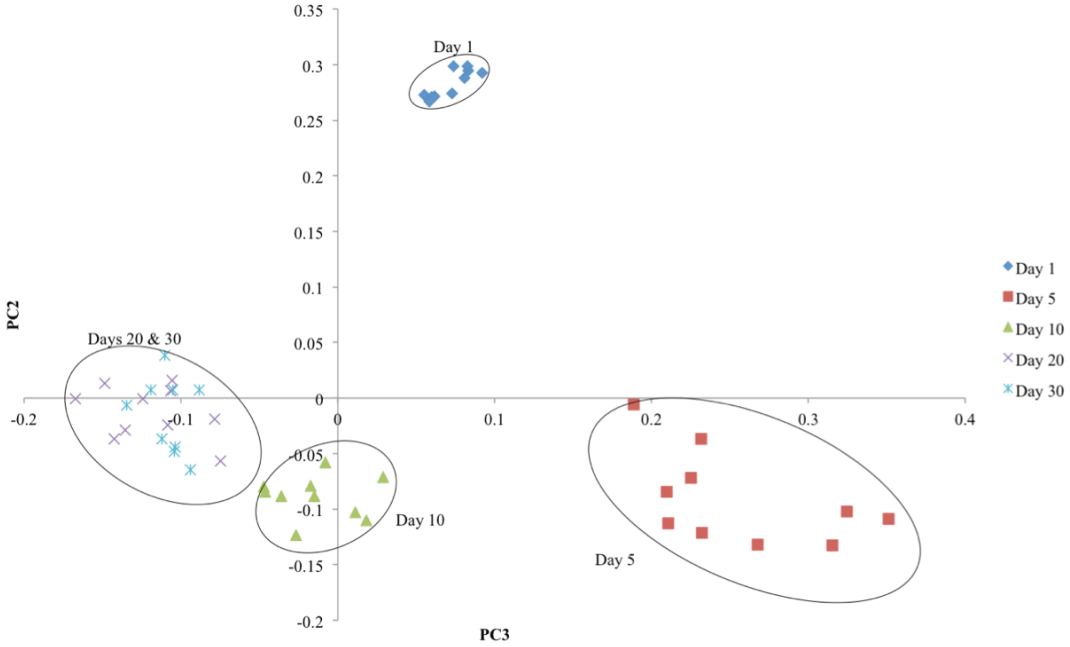

Figure 5: PCA plot showing PC3 against PC2 for *C. vicina* with clustering days circled

ACCEPTED MANUSCRIPT

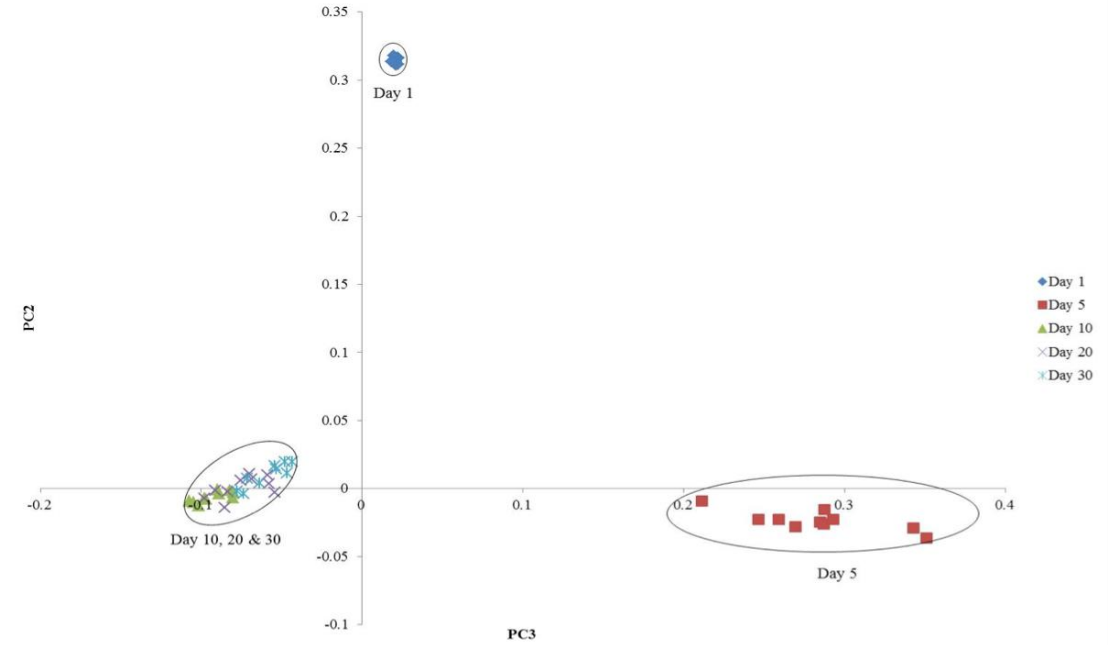

Figure 6: PCA plot showing PC3 against PC2 for *C. vomitoria* with clustering days circled

ACCEPTED MANUSCRIPT

Table 1: List of the compounds extracted and used for subsequent PCA from the adult flies of *L. sericata*, along with the total percentage of each compound present, the percentage standard deviation for each day and the Kovats Index to aid identification

			Day 1	**Day 5**	**Day 10**
Peak	**Peak**	**Kovats**	***n*=10**	***n*=10**	***n*=10**
no.	**Identification**	**iu**	**%**	**%**	**%**
1	Henicosane	2100	tr	0.72±0.32	0.83±0.48
2	Tricosene[2]	2265	tr	tr	1.23±0.94
3	Tricosane	2300	0.63±0.18	11.62±6.87	19.21±17.86
4	9+11-Methyltricosane	2337	tr	0.48±0.23	tr
5	Tetracosane	2400	tr	tr	0.69±0.37
6	2-Methyltetracosane	2464	tr	0.64±0.22	0.65±0.30
7	Pentacosene[2]	2472	tr	tr	2.15±1.86
8	Pentacosane	2500	2.28±0.44	8.20±4.89	16.06±9.20
9	9+11-Methylpentacosane	2539	0.81±0.24	1.43±0.67	1.58±0.89
10	7-Methylpentacosane	2544	0.84±0.25	tr	tr
11	3-Methylpentacosane	2574	0.67±0.30	tr	tr
12	Hexacosane	2600	0.69±0.13	0.66±0.44	0.93±0.39
13	8-Methylhexacosane	2611	0.68±0.27	tr	tr
14	2-Methylhexacosane	2664	tr	1.60±0.57	1.78±0.89
15	Heptacosene[2]	2671	tr	0.49±0.46	4.71±3.88
16	Heptacosene[2]	2678	0.78±0.36	7.53±4.11	1.53±0.65
17	Heptacosene[2]	2689	tr	0.31±0.17	tr
18	Heptacosane	2700	8.60±1.41	13.27±6.15	18.36±7.22
19	13-Methylheptacosane	2734	3.57±0.68	2.08±1.40	1.51±0.54
20	9-Methylheptacosane	2738	1.46±0.32	tr	tr
21	7-Methylheptacosane	2743	1.26±0.25	tr	tr
22	5-Methylheptacosane	2752	1.35±0.29	tr	tr
23	2-Methylheptacosane	2769	0.70±0.22	tr	tr
24	3-Methylheptacosane	2775	7.66±1.40	1.85±0.75	0.62±0.24
25	Octacosane	2800	1.03±0.17	0.80±0.30	0.64±0.21
26	12+14-Methyloctacosane	2841	2.01±0.36	0.71±0.36	tr
27	8-Methyloctacosane	2845	0.55±0.12	tr	tr
28	2-Methyloctacosane	2872	5.29±0.89	8.86±3.15	5.01±1.68

ACCEPTED MANUSCRIPT

29	Nonacosene[2]	2879	2.63±0.78	5.51±3.06	7.85±4.57
30	Nonacosene[2]	2885	1.57±0.45	6.61±3.32	1.55±0.31
31	2,12/2,14-Dimethyloctacosane[1]	2897	1.04±0.18	0.73±0.50	tr
32	Nonacosane	2900	4.08±0.89	4.37±1.70	4.10±1.84
33	11+15-Methylnonacosane	2939	19.23±2.94	10.88±5.31	2.88±1.02
34	9-Methylnonacosane	2943	6.25±5.02	0.54±0.24	tr
35	7-Methylnonacosane	2948	3.78±0.55	tr	tr
36	5-Methylnonacosane	2957	1.10±0.19	tr	tr
37	9,17-Dimethylnonacosane[1]	2972	2.26±0.36	tr	tr
38	3-Methylnonacosane	2978	1.80±0.28	0.65±0.28	tr
39	12+14+16+18-Methyltriacontane	3035	1.81±0.33	tr	tr
40	Hentriacontadiene[2]	3064	tr	0.97±0.66	tr
41	2-Methyltriacontane	3068	5.09±0.95	3.62±1.28	2.32±0.64
42	Hentriacontene[2]	3078	tr	1.43±0.55	1.20±0.40
43	2,12/2,14-Dimethyltricontane[1]	3096	1.88±0.56	0.73±0.32	0.60±0.17
44	2,8-Dimethyltriacontane[1]	3099	0.56±0.15	tr	tr
45	11+13+15-Methylhentriacontane	3121	3.05±0.56	1.48±0.93	1.05±0.31
46	9-Methylhentriacontane	3123	1.02±0.23	tr	tr
47	11,19-Dimethylhentriacontane[1]	3143	1.98±0.39	0.61±0.30	tr
48	Tritriacontene[2]	3217	tr	0.60±0.28	0.93±0.28

[1] Tentative identification based on Kovats Index

[2] Double bond positions determined for adult flies (see page 126)

tr = Trace amount detected (<0.5%)

ACCEPTED MANUSCRIPT

Table 2: List of the compounds extracted and used for subsequent PCA analysis from the adult flies of *C. vicina*, along with the total percentage of each compound present, the percentage standard deviation for each day and the Kovats Index to aid identification

			Day 1	Day 5	Day 10	Day 20	Day 30
Peak	**Peak**	**Kovats**	***n*=10**	***n*=10**	***n*=10**	***n*=10**	***n*=10**
no.	**Identification**	**iu**	**%**	**%**	**%**	**%**	**%**
1	Heneicosane	2100	tr	0.66±0.23	tr	tr	tr
2	2-Methyldocosane	2269	tr	3.84±3.63	1.99±0.91	0.82±0.38	1.97±1.12
3	Tricosene[1]	2273	tr	4.63±3.48	2.57±0.57	0.63±0.20	tr
4	Tricosene[1]	2275	tr	0.79±0.53	tr	tr	tr
5	Tricosane	2300	0.42±0.19	9.10±3.99	5.47±0.81	2.68±0.73	tr
6	11+9-Methyltricosane	2338	tr	1.22±0.77	0.91±0.26	tr	tr
7	7-Methyltricosane	2343	tr	0.89±0.78	tr	tr	tr
8	3-Methyltricosane	2372	tr	0.61±0.35	tr	tr	tr
9	2-Methyltetracosane + Pentacosene[1]	2446	tr	11.30±6.01	1.17±0.43	tr	tr
10	Pentacosene[1]	2475	tr	3.12±1.53	10.72±4.96	4.35±1.14	5.73±4.67
11	Pentacosene[1]	2476	tr	0.93±0.51	2.31±0.64	1.72±0.61	2.89±2.41
12	Pentacosene[1]	2484	tr	0.82±0.27	2.31±0.81	1.29±0.47	1.18±1.04
13	Pentacosane	2500	3.52±0.57	12.53±5.13	17.60±4.40	18.74±7.08	21.12±11.53
14	11+9-Methylpentacosane	2535	1.43±0.44	4.80±2.54	2.98±1.36	2.72±0.85	3.39±1.76
15	7-Methylpentacosane	2543	0.74±0.22	tr	tr	tr	tr
16	3-Methylpentacosane	2574	1.17±0.34	1.26±0.60	0.94±0.46	0.70±0.35	0.91±0.68
17	Hexacosane	2600	1.14±0.44	0.80±0.22	0.86±0.27	0.87±0.31	0.83±0.33
18	3,7-Dimethylpentacosane[2]	2610	1.09±0.35	tr	tr	tr	tr
19	12+14+16-Methylhexacosane	2633	0.83±0.30	tr	tr	tr	tr
20	Heptacosadiene	2655	tr	tr	1.81±0.7	1.54±0.42	1.15±0.69

ACCEPTED MANUSCRIPT

					2		
21	Heptacosene[1*]	2661	3.49±0.78	12.47±5.49	11.37±8.20	15.31±16.58	17.57±16.20
22	Heptacosene[1]	2667	tr	2.25±0.76	3.46±1.48	2.49±0.70	2.38±1.34
23	Heptacosene[1]	2671	tr	2.25±0.76	4.09±2.72	5.83±3.48	4.26±3.27
24	Heptacosane	2700	15.12±2.72	10.67±4.92	14.48±3.85	22.74±9.70	21.59±8.31
25	11+13-Methylheptacosane	2734	9.32±2.27	2.68±1.86	1.18±0.84	0.95±0.38	1.18±0.49
26	9-Methylheptacosane	2737	1.95±0.43	tr	tr	tr	tr
27	7-Methylheptacosane	2742	2.15±0.47	tr	tr	tr	tr
28	5-Methylheptacosane	2750	1.89±0.41	tr	tr	tr	tr
29	9,13+9,15-Dimethylheptacosane[2]	2764	1.33±1.17	tr	tr	tr	tr
30	3-Methylheptacosane	2776	6.28±1.21	1.13±0.55	0.72±0.36	tr	tr
31	5,9+5,13-Dimethylheptacosane[2]	2784	2.53±0.57	tr	tr	tr	tr
32	Octacosane	2800	1.96±0.46	1.76±3.32	0.76±0.35	tr	tr
33	3,9+3,11-Dimethylheptacosane[1]	2811	0.90±0.29	tr	tr	tr	tr
34	Trimethylheptacosane[2]	2815	2.47±0.54	tr	tr	tr	tr
35	12+14+16-Methyloctacosane	2840	2.39±0.53	tr	tr	tr	tr
36	3,7,15+3,7,15-Trimethylheptacosane[2]	2847	1.11±0.32	tr	tr	tr	tr
37	6-Methyloctacosane	2853	0.76±0.17	tr	tr	tr	tr
38	Nonacosadiene	2860	tr	0.80±0.43	2.83±1.49	4.70±1.40	2.94±2.26
39	Nonacosadiene	2865	tr	tr	0.92±0.45	1.48±0.81	1.41±1.28
40	4-Methyloctacosane	2867	4.71±1.80	tr	tr	tr	tr
41	2-	2871	1.28±1.41	3.36±1.82	3.08±2.15	2.69±1.43	3.04±2.26

ACCEPTED MANUSCRIPT

	Methyloctacosane						
42	Nonacosene[1]	2880	tr	1.22±0.80	2.37±1.30	2.64±1.14	2.31±2.04
43	Nonacosene[1]	2887	tr	tr	tr	1.81±2.80	0.72±1.18
44	6,14-Dimethyloctacosane[2]	2882	2.53±0.57	tr	tr	tr	tr
45	2,12+2,14-Dimethyloctacosane[2]	2895	1.44±0.78	tr	tr	tr	tr
46	Nonacosane	2900	5.47±1.30	2.30±1.55	1.91±1.08	2.11±0.75	2.21±1.01
47	x,6-Dimethyloctacosane + x,10,14-Trimethyloctacosane	2904	1.13±0.31	tr	tr	tr	tr
48	4,8,12+4,8,14-trimethyloctacosane[2]	2921	1.34±0.39	tr	tr	tr	tr
49	9+11+1315-Methylnonacosane	2941	8.25±1.92	0.90±0.52	tr	tr	tr
50	7-Methylnonacosane	2949	1.98±0.47	tr	tr	tr	tr
51	5-Methylnonacosane	2957	1.16±0.42	tr	tr	tr	tr
52	11,15-Dimethylnonacosane[2]	2966	0.63±0.20	tr	tr	tr	tr
53	9,17-Dimethylnonacosane[2]	2972	0.89±0.24	tr	tr	tr	tr
54	3-Methylnonacosane +7,11-Dimethylnonacosane[2]	2979	2.45±0.54	tr	tr	tr	tr
55	trimethylnonacosane	3015	0.65±0.13	tr	tr	tr	tr
56	2-Methyltriacontane	3070	1.07±0.29	0.90±0.33	1.18±0.52	1.19±0.44	1.23±0.56
57	11+13-Methylhentriacontane	3124	1.05±0.28	tr	tr	tr	tr

[1] Double bond positions determined – alkadiene bond position not determined

[2]Tentative identification based on Kovats Index values and match with NIST08 Library database

ACCEPTED MANUSCRIPT

*2-MeC26:H in day 1 then co-elutes with C27:1 in day 5 onwards

Tr = Trace amount detected (<0.5%)

ACCEPTED MANUSCRIPT

Table 3: List of the compounds extracted and used for subsequent PCA from the adult flies of *C. vomitoria*, along with the total percentage of each compound present, the percentage standard deviation for each day and the Kovats Index to aid identification

			Day 1	**Day 5**	**Day 10**	**Day 20**	**Day 30**
Peak no.	**Peak Identification**	**Kovats iu**	***n*=10** **%**	***n*=10** **%**	***n*=10** **%**	***n*=10** **%**	***n*=10** **%**
1	Heneicosane	2100	tr	3.44±1.26	1.89±1.53	0.72±0.21	tr
2	2-Methylhenicosane	2272	tr	31.64±8.45	1.01±0.44	1.37±0.68	0.95±0.62
3	Tricosene[1]	2276	tr	2.99±1.59	30.50±12.99	18.80±3.76	9.67±6.31
4	Tricosene[1]	2275	tr	tr	tr	tr	4.02±2.22
5	Tricosane	2300	3.24±2.41	26.10±7.74	38.44±18.03	33.95±9.37	25.51±10.37
6	9+11-Methyltricosane	2337	tr	1.22±0.42	tr	tr	0.77±0.23
7	7-Methyltricosane	2342	tr	0.76±0.29	tr	tr	tr
8	Tetracosene[1] + 3-Methyltricosane	2373	tr	0.75±0.26	tr	tr	0.62±0.24
9	Tetracosane	2400	tr	tr	tr	tr	0.70±0.17
10	2-Methyltetracosane	2466	0.88±0.85	2.00±1.77	2.14±1.19	3.62±1.11	2.94±1.23
11	Pentacosene[1]	2472	1.47±1.26	9.53±4.00	11.79±5.30	14.66±2.53	12.50±6.48
12	Pentacosene[1]	2476	tr	tr	tr	4.92±1.17	7.14±3.08
13	Pentacosane	2500	3.04±1.51	8.80±1.99	9.23±4.02	15.87±3.67	18.19±5.16
14	9+11-Methylpentacosane	2536	2.67±1.91	1.83±0.78	tr	1.00±0.20	2.40±0.63
15	2-Methylpentacosane	2664	2.21±1.64	0.68±0.29	tr	0.65±0.24	1.45±0.45

ACCEPTED MANUSCRIPT

16	Heptacosene[1]	2671	0.86±0.54	2.61±1.74	1.72±1.09	1.68±0.64	3.61±1.87
17	Heptacosene[1]	2675	tr	tr	tr	tr	1.54±0.80
18	Heptacosane	2700	5.83±2.81	2.76±0.91	1.83±0.85	1.89±0.68	3.76±1.52
19	11+13-Methylheptacosane	2734	3.40±2.20	0.53±0.25	tr	tr	0.63±0.18
20	9-Methylheptacosane	2738	0.91±0.48	tr	tr	tr	tr
21	5-Methylheptacosane	2753	0.83±0.39	tr	tr	tr	tr
22	9,15+9,17-Dimethyheptacosane[2]	2765	0.91±0.48	tr	tr	tr	tr
23	3-Methylheptacosane	2775	3.27±1.12	tr	tr	tr	tr
24	Octacosane	2800	0.83±0.39	tr	tr	tr	tr
25	2-Methyloctacosane	2868	5.98±4.28	0.57±0.18	tr	tr	0.99±0.40
26	Nonacosene[1]	2879	2.64±0.68	0.61±0.22	tr	tr	0.65±0.34
27	Nonacosene[1]	2885	0.98±0.36	tr	tr	tr	tr
28	Nonacosane	2900	8.16±4.25	1.71±0.71	1.45±0.92	0.87±0.22	1.37±0.55
29	2, 6-Dimethyloctacosane[3]	2905	1.35±0.76	tr	tr	tr	tr
30	9+11-Methylnonacosane	2938	8.84±3.66	tr	tr	tr	tr
31	7-Methylnonacosane	2947	2.96±0.96	tr	tr	tr	tr
32	5-Methylnonacosane	2957	0.91±0.32	tr	tr	tr	tr
33	^{x}9,x+^{y}11,x-Dimethylnonacosane[4]	2966	0.83±0.40	tr	tr	tr	tr
34	7,x-Dimethylnona	2973	3.07±1.31	tr	tr	tr	tr

ACCEPTED MANUSCRIPT

	cosane[2] and 3-Methylnonacosane						
35	5,13+5,15+5,17-Dimethylnonacosane[2]	2986	1.07± 0.61	tr	tr	tr	tr
36	3,7-Dimethylnonacosane[3]	3013	1.18± 0.44	tr	tr	tr	tr
37	12+14+16-Methyltriacontane	3036	0.86± 0.39	tr	tr	tr	tr
38	8-Methyltriacontane	3044	0.59± 0.25	tr	tr	tr	tr
39	2-Methyltriacontane	3070	3.19± 1.68	tr	tr	tr	0.60±0.28
40	6,14-Dimethyltriacontane[2]	3081	1.41± 0.41	tr	tr	tr	tr
41	2,12+2,14-Dimethyltriacontane[3]	3099	2.55± 1.26	tr	tr	tr	tr
42	Hentriacontane	3100	tr	0.77± 0.54	tr	tr	tr
43	2,6+2,8-Dimethyltriacontane[3]	3103	3.45± 1.87	tr	tr	tr	tr
44	4,8,14-Trimethyltriacontane[2]	3116	1.60± 0.77	tr	tr	tr	tr
45	2,6,14-Trimethyltriacontane[2]	3125	7.40± 3.17	0.72± 0.33	tr	tr	tr
46	2,6,10,14-Tetramethyltricontane[2]	3141	1.95± 1.01	tr	tr	tr	tr
47	7,15-Dimethylhentricaontane[2]	3148	2.39± 1.01	tr	tr	tr	tr
48	5,15-Dimethylhentriacontane[2]	3156	0.58± 0.20	tr	tr	tr	tr
49	Unknown Hydrocarbon	3165	0.50± 0.13	tr	tr	tr	tr
50	5,9,13+5,11,15-	3174	1.06± 0.38	tr	tr	tr	tr

ACCEPTED MANUSCRIPT

	triMethylhentriacontane[2]						
51	Unknown Hydrocarbon	3209	0.71± 0.22	tr	tr	tr	tr
52	17-Methyltritriacontane	3229	0.92± 0.49	tr	tr	tr	tr
53	7,17-Dimethyltritriacontane[2]	3255	1.22± 0.54	tr	tr	tr	tr
54	7,11,15-Trimethyltritriacontane[2]	3272	1.28± 0.63	tr	tr	tr	tr

[1]Double bond position determined (see page 126). Bond position not determined for alkadienes

[2]Tentative identification based on Kovats Index values and match with NIST08 Library database

[3]Position of the first methyl was established by using the Kovats Index [19]

[4]x = 15, 17, 19

y = 15, 17

tr = Trace amount detected (<0.5%)

ACCEPTED MANUSCRIPT

Table 4: The overall test performance of each SOM when classifying the adult of *L. sericata*, *C. vicina* and *C. vomitoria* hydrocarbon profiles with the SOM output layer that provided the best performance shown in brackets after the species name.

Test approach	**% correct (SD)**		
	***L. sericata* (6x6)**	***C. vicina* (13x13)**	***C. vomitoria* (10x10)**
Average of five samples	100 (0)	88 (19.39)	98 (4)
Individual samples	93.33 (5.73)	70.6 (15.59)	85.4 (12.73)

ACCEPTED MANUSCRIPT

Table 5: Confusion matrices showing the performance of each SOM when classifying for each fold of cross-validation as well as the overall classification performance for each day when tested using the average of the remaining five input patterns of *L. sericata* (top), *C. vicina* (middle) and *C. vomitoria* (bottom) hydrocarbon profiles.

SOM classification	Input pattern tested		
	D1	D5	D10
D1	10		
D5		10	
D10			10
% correct	100	100	100

SOM classification	Input pattern tested				
	D1	D5	D10	D20	D30
D1	10				
D5		9			
D10			10		
D20				5	
D30		1		5	10
% correct	100	90	100	50	100

SOM classification	Input pattern tested				
	D1	D5	D10	D20	D30
D1	10				
D5		10			
D10			9		
D20			1	10	
D30					10
% correct	100	100	90	100	100